AL/AO-TR-1995-0121

DTIC ELECTE NOV 0 7 1995 S D B

A R M S T R O N G L A B O R A T O R Y

COMPARISON OF A COMPUTERIZED VERSION TO A PAPER/PENCIL VERSION OF THE MULTIDIMENSIONAL APTITUDE BATTERY

Paul D. Retzlaff
Raymond E. King
Joseph D. Callister

AEROSPACE MEDICINE DIRECTORATE
CLINICAL SCIENCES DIVISION
NEUROPSYCHIATRY BRANCH
2507 Kennedy Circle
Brooks Air Force Base, TX 78235-5117

July 1995

Final Technical Report for Period March 1994 – July 1995

Approved for public release; distribution is unlimited.

19951106 025

DTIC QUALITY INSPECTED 5

AIR FORCE MATERIEL COMMAND
BROOKS AIR FORCE BASE, TEXAS

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this technical report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

Raymond E. King

RAYMOND E. KING, Major, USAF, BSC
Project Scientist

JOE EDWARD BURTON, Colonel, USAF, MC, CFS
Chief, Clinical Sciences Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1995	Final Technical Report Mar 94 - Jul 95

4. TITLE AND SUBTITLE
Comparison of a Computerized Version to a Paper/Pencil Version of the Multidimensional Aptitude Battery (MAB)

5. FUNDING NUMBERS
Proj: ILIR
Task: AC
WU: 41

6. AUTHOR(S)
Retzlaff, Paul D.
King, Raymond E.
Callister, Joseph D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Armstrong Laboratory (AFMC)
Aerospace Medicine Directorate
Clinical Sciences Division, Neuropsychiatry Branch
2507 Kennedy Circle
Brooks AFB TX 78235-5117

8. PERFORMING ORGANIZATION
AL/AO-TR-1995-0121

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING

11. SUPPLEMENTARY NOTES
Armstrong Laboratory Technical Monitor: Major Raymond E. King,
(210) 536-3537

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This study examined the comparability of the Armstrong Laboratory's computerized version and the original paper-and-pencil version of an intelligence test. The Multidimensional Aptitude Battery (MAB) is a multiscale test of intelligence that is widely used in aerospace cognitive testing. The research question was whether the two tests are psychometrically equivalent. Comparing the scores of 135 student pilot candidates who took the paper-and-pencil version to the scores of 402 student pilot candidates who took the computerized version, there are no clinically significant differences between the two versions. Full Scale, Verbal, and Performance Intelligence Quotient (IQ) scores were not significantly different across the two tests. Single factor and two factor analyses indicated that the computerized version was factorially similar to not only the paper-and-pencil pilot candidate data but also the original construction samples.

14. SUBJECT TERMS
IQ
Computerized Testing

15. NUMBER OF PAGES
14

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89) Prescribed by ANSI Std Z-39-18
298-102 COMPUTER GENERATED

CONTENTS

TABLES

FIGURES

ssion For
GRA&I ☑
TAB ☐
ounced ☐
fication
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

PREFACE

This project was completed under ILIRAC41 in support of the Multidimensional Aptitude Battery computerization project. Funding for this and other projects is through Armstrong Laboratory and the Air Force Medical Operating Agency.

Appreciation is extended to the technical support staff of the project including TSgt. Dwayne C. Lanier, SSgt. Pauline Etterle, Leonard Longo, William D. Taylor and William M. Weaver. Additionally, Malcolm J. Ree and Christopher F. Flynn are thanked for seeing the need for this project and preparing an earlier version of the proposal.

COMPARISON OF A COMPUTERIZED VERSION TO A PAPER/ PENCIL VERSION OF THE MULTIDIMENSIONAL APTITUDE BATTERY (MAB)

SUMMARY

This study examined the comparability of the Armstrong Laboratory's computerized version and the original paper-and-pencil version of an intelligence test. The Multidimensional Aptitude Battery (MAB) is a multiscale test of intelligence that is widely used in aerospace cognitive testing. The research question was whether the two tests are psychometrically equivalent. Comparing the scores of 135 student pilot candidates who took the paper-and-pencil version to the scores of 402 student pilot candidates who took the computerized version, there are no clinically significant differences between the two versions. Full Scale, Verbal, and Performance IQ scores were not significantly different across the two tests. Single factor and two factor analyses indicated that the computerized version was factorially similar to not only the paper-and-pencil pilot candidate data but also the original construction samples. Further, internal consistencies are higher for the computerized version than for the paper-and-pencil version for pilot candidate data. Finally, visual analysis of the distributions suggests no major differences.

INTRODUCTION

Background

Clinical psychologists from the Neuropsychiatry Branch of Armstrong Laboratory accomplish and supervise the psychological assessment of many aviators each year. A number of these assessments are completed in a conventional manner such as when an aviator requests a medical waiver in order to return to flying status. However, the majority of psychological assessments are completed in less conventional ways. For example, over 1,000 student pilot candidates are evaluated as part of the Enhanced Flight Screening (EFS) Program each year. These students are medically screened in groups of 20-24 prior to beginning flight screening at Hondo, TX, or they are medically screened in groups of 6-8 before beginning flight screening at the Air Force Academy. All psychological testing must be completed in one 4-hour block of time. Time constraints and the need for group administration has motivated the clinical psychologists at Armstrong Laboratory to identify and develop more efficient ways of administering and scoring psychological tests.

The Multidimensional Aptitude Battery (MAB) was developed as a measure of intelligence similar to the Wechsler Adult Intelligence Scale- Revised (WAIS-R) but permits group administration, automated administration, and hand/ machine scoring (Jackson, 1984). The MAB provides summary Full Scale, Performance and Verbal IQ Scores as well as subtest scores.

Subtest and summary scores of the MAB and the WAIS-R correlate to about the same degree as WAIS-R scores correlate with the original version WAIS intelligence test scores. The MAB was first used with aviators in the 1980's (Retzlaff and Gibertini, 1987; 1988).

In 1994, Flynn, Sipes, Grosenbach, & Ellsworth completed a study of rated pilots using a partially computerized version of the MAB developed by the test publisher. Although this version was more efficient than previous versions, it still required the use of a printed stimulus booklet. Therefore, a fully computerized version was developed by the Neuropsychiatry Branch of Armstrong Laboratory, in co-operation with the test's author, Douglas Jackson, Ph.D. This fully computerized version is currently being used in the evaluation of all EFS students (King and Flynn, 1995) and in a Defense Women's Health Initiative Program study entitled "Assessment of Psychological Factors in Aviators." The purpose of the present study was to determine whether the Armstrong Laboratory's computerized version of the MAB was psychometrically equivalent to the original paper-and-pencil version.

Method Review

There are three approaches to the comparison of two versions of a single test. The first and second approaches involve the comparison of the two tests taken by a single group. The third would look at differences across two sample, each sample taking only one version of the test.

The first approach would be considered a classic alternate forms study. Here a large number of subjects, perhaps 100 to 200 would take both forms of a test. Level of difficulty would be assessed by comparing the two sets of scores for significant differences. It may be that one version is "easier" than the other and results in artificially higher scores. The second indicator would look at the correlations of the two versions of the test. This, in essence, determines if subjects are relatively positioned on the two score distributions at similar points. If the two versions are similar and truly alternate forms the correlations between the test scores should be positive and high. These coefficients should approach reliability and, therefore, be in the high 0.80's. The difficulty with this approach is that it requires a great deal of testing time over a period of weeks and is often impractical.

The second approach is invoked based upon the findings of the first. If significant differences in difficulty or distribution are found, then it is often of value to "equate" the two score distributions. Here the two scores are often transformed into a third score which is optimized to the two underlying scores. Here, this third score can be used to examine scores regardless of the test form from which they are derived. Carretta and Ree (1993) at Armstrong Laboratory have used this

technique with great success. This approach, however, is only necessary when a large set of existing testing is available and needs to be made comparable to a newer/ different test version. This also requires very large number of subjects to map each percentage point of performance. Sample sizes of 1000 would often be needed.

The third approach would look at the relative difficulty of the two versions and the variability of the distributions across two different samples. Here one large sample taking one version of the test would be compared to another sample taking the other version. Samples of 100 to 200 would be adequate. Significant differences in level of difficulty could be determined statistically. Additionally, the degree of variance in the scores could be compared as well as other parameters of distribution shape. Secondary issues of reliability and factorial stability can also be examined. Alternate forms validity coefficients cannot be calculated, unfortunately. This approach is often the most practical.

Purpose

The purpose of the current study is to determine if the two versions of the MAB are similar. Questions include 1) whether or not IQ scores are of similar level and variability across the two versions, 2) whether the performance subtests survived the transition well, and 3) whether the verbal subtests are behaving as expected.

METHOD

Subjects

Two Air Force samples participated in this study. The first was a group of 135 student pilot candidates and the second was a group of 402. The sample as a whole had a mean age of 23.5 (sd 4.2) and 5% were female. Subjects who had been commissioned through Officer Training School, ROTC, and the Air National Guard were all college graduates. Approximately, 42% were Juniors at the United States Air Force Academy. There were no significant differences between the groups on demographic variables.

Measures

Two versions of the MAB were used. The first version was the original designed by Jackson (1984). It is a paper-and-pencil version. There are 10 subtests each with a time limit of 7 minutes. Subjects read items from a booklet and endorse a, b, c, or d on a bubble sheet. The bubble sheets can be hand scored, computer scored locally, or mailed to the test company for computer scoring. They were computer scored locally for this study.

The second version is the Armstrong Laboratory's computerized version. Here verbal type questions are presented as text on a computer screen and subjects are asked to respond to the computer with an a, b, c, or d response with a light pen or keyboard entry. The performance type items were scanned into computer graphic files and are presented in a window on the monitor. This computerization was done and is used with the consent of the test author with explicit copyright permission.

Procedures

Prior to entering the Enhanced Flight Screening programs at Hondo, TX, and the Air Force Academy in Colorado Springs, CO, student pilot candidates are asked to participate in baseline cognitive testing. They are additionally asked, but not required, to participate in personality testing.

Students tested on the original paper-and-pencil version were administered the test in accordance with the procedures outlined in the manual. Booklets are handed out and a proctor ensures that all subjects in a group are given the appropriate 7 minutes per subtest.

Students tested on the computer were given items and timed by the computer. The subtests would begin and end in accordance with the programming of the batch files. While groups of students were tested simultaneously, testing in this manner is more individual in nature.

Analysis

The data were analyzed for differences across testing conditions. Mean levels of performance were analyzed. Underlying subtest intelligence loadings were calculated. The concordance of 2 factor structures across conditions were compared to the construction sample. Finally, the internal consistency of the Full Scale IQ scores were calculated and compared.

RESULTS

Differences in mean levels

Table 1 presents the means and standard deviations for the 10 subtests as well as the Full, Verbal, and Performance IQ summary scores. Subtest data are raw scores. This was done to avoid any score changes as a function of scale score conversion. Summary IQ scores are in scaled format with the usual mean of 100 and standard deviation of 15 for the population at large.

The mean Full Scale IQ for the students who took the paper

and pencil version was 120 and the mean for the students taking the computer version was 119. This is not a significant difference ($\underline{t}$=1.35, df=535, $\underline{p}$=.1761). Further, no differences between groups on the Verbal IQ was found with means of 119 and 118, respectively. Performance IQ's were 119 and 118, respectively, again, representing no significant differences.

Table 1

Means, standard deviations, and t-tests for all MAB variables.

Variable	Paper	Computer	$\underline{t}$	$\underline{p}$
Full Scale	120.1 (6.6)	119.1 (7.1)	1.35	.1761
Verbal	118.5 (6.9)	117.9 (7.1)	0.90	.3697
Performance	119.0 (8.3)	117.7 (8.9)	1.48	.1400
Information	29.8 (4.0)	29.3 (4.7)	1.10	.2696
Comprehension	23.3 (2.1)	23.4 (2.2)	-.26	.7934
Arithmetic	15.7 (2.2)	15.6 (2.0)	0.37	.7142
Similarities	27.6 (3.0)	27.8 (3.0)	-.63	.5289
Vocabulary	28.8 (5.5)	29.3 (5.8)	-.74	.4607
Digit Symbol	28.1 (3.6)	29.6 (3.2)	-4.74	.0001*
Picture Completion	26.9 (3.7)	26.9 (3.7)	0.02	.9826
Spatial	37.4 (6.3)	36.6 (6.9)	1.09	.2765
Picture Arrangement	13.8 (2.0)	12.3 (2.0)	7.81	.0001*
Object Assembly	15.9 (3.2)	15.7 (3.1)	0.51	.6129

Note: Summary IQ scores are in scaled score units. Subtest data is in raw score units. * denotes significant differences.

With respect to the variance of the scores, little difference is seen across groups. The standard deviations for the Full Scale IQ are both about 7. The standard deviations for the Verbal IQ scores are also both about 7 and the Performance standard deviations are about 8 and a half.

No differences between means were found for any of the Verbal subtests. These subtests included Information, Comprehension, Arithmetic, Similarities, and Vocabulary. Again few if any differences are noticeable in the standard deviations. The largest difference is on the Information subscale with the paper version students having a standard deviation of 4.0 and the computer students having a 4.7. This ratio is only 1.18.

No mean differences for Performance subtests were found for the Picture Completion, Spatial, or Object Assembly tasks. There was a significant difference between scores on the Digit Symbol subtest. Here the computer version resulted in a mean score of

29.6 and the paper version a 28.1. In scaled score units (mean of 50 and standard deviation of 10), this is a 67 versus a 64. Additionally, there was a significant mean difference between groups on Picture Arrangement. Here the paper version resulted in a higher score of 13.8 than the computer version with a 12.3. Scaled scores would be 65 and 60.

Variances again appear equal across Performance subtests. The greatest difference would be on the Digit Symbol with standard deviations of 3.6 and 3.2. Here the ratio would be 1.12.

In summary, few if any differences are found for level of performance. Additionally, no significant differences were found for the important three summary IQ scores.

Single factor solution/ "g" loadings

In order to assess the degree to which the subscales of the two versions of the test correlate with a singular general intelligence dimension, a factor analysis was done which extracted and rotated only one factor. This was compared across samples as well as to the loadings presented in the manual from the construction sample.

Table 2

Single factor subtest loadings for Jackson normative sample, pilot paper version, and pilot computer version.

Variable	Jackson	Paper	Computer
Information	.77	.71	.70
Comprehension	.82	.55	.62
Arithmetic	.68	.37	.49
Similarities	.79	.67	.69
Vocabulary	.73	.61	.69
Digit Symbol	.53	.30	.46
Picture Completion	.67	.42	.64
Spatial	.56	.59	.55
Picture Arrangement	.63	.42	.57
Object Assembly	.65	.54	.62
Variance	N/R	28%	37%
N	3121	135	402

Table 2 presents the three vectors. The Jackson sample

generated a solution with generally high loadings (scale- factor correlations). Verbal subtests load higher than Performance subtests. The lowest loading is for Digit Symbol.

The paper-and-pencil version given to student pilots resulted in generally lower loadings. Again, the Verbal subtests load higher than Performance subtests. Again, Digit Symbol has the lowest loading. This solution resulted in only 28% of the variance being modeled.

The computerized version resulted in loadings that were higher than the loadings from the pilot paper-and-pencil version. They were not though as high as the Jackson sample with its 3121 size sample. Digit Symbol is again the lowest. Interestingly, the computer version data models far more variance at 37% than the pilot paper-and-pencil version.

In summary, the loadings of the three samples on a single intelligence factor are supportive of the two versions being similar. If anything, these data suggest that the computerized version is superior to the paper-and-pencil version for pilots.

Two factor solution

In order to determine if the subscales group logically into verbal and performance factors, a two factor solution for the two pilot samples was compared to the Jackson construction sample.

Table 3

Two factor solution for Jackson normative sample, pilot paper version, and pilot computer version.

Variable	Factor 1			Factor 2		
	J	P	C	J	P	C
Information	.83*	.81*	.79*	.23	.14	.15
Comprehension	.83*	.67*	.67*	.28	.06	.18
Arithmetic	.54*	.18	.34*	.43*	.35*	.36*
Similarities	.81*	.62*	.76*	.25	.31*	.17
Vocabulary	.82*	.77*	.83*	.14	.05	.09
Digit Symbol	.17	-.18	.05	.63*	.66*	.65*
Picture Completion	.44*	.24	.44*	.53*	.35*	.48*
Spatial	.10	.05	.14	.77*	.84*	.68*
Picture Arrangement	.30*	.16	.16	.63*	.45*	.69*
Object Assembly	.22	.18	.17	.75*	.61*	.75*
Variance	N/R	23%	27%	N/R	21%	24%
N	3121	135	402	same		

Note: J heading means Jackson data, P heading pilot paper version, and C heading pilot computer version. * denotes loadings of .30 and greater. N/R denotes that the variance accounted for percentage was not reported for the Jackson data in the test manual.

It is hoped that the five Verbal subscales will form a common factor as will the five Performance subscales. As can be seen in Table 3, all three samples generally result in a clean and logical two factor solution.

Factor 1 represents the five Verbal subtests. Here the Jackson data provides high loadings for all but the Arithmetic subtest. Interestingly, the pilot computer sample has generally higher loadings than the paper-and-pencil pilot sample. Both pilot samples also exhibit the lowest loading for Arithmetic as does the Jackson sample.

Factor 2 represents the performance factor. All three samples display loadings which are similar. Picture Completion has the lowest loading for all three samples. Again, remarkable concordance across patterns is seen.

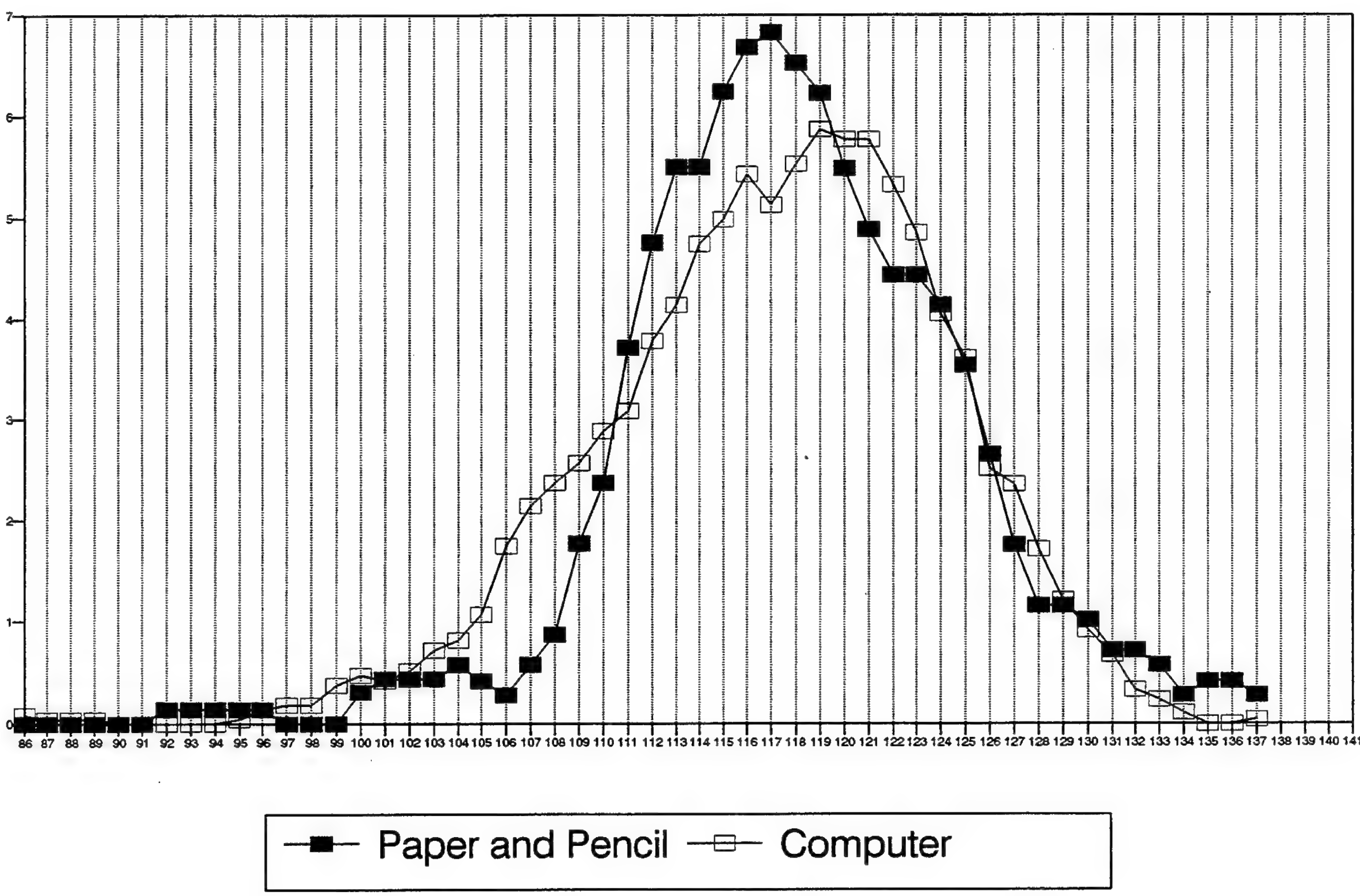
The factors modeled 44% of the paper-and-pencil pilot sample data and 51% of the computer pilot data.

In summary, the two factor solutions are more alike than they are different and represent excellent factor concordance. Again, an argument can be made that for pilots the computerized version behaves better than the paper-and-pencil version.

Internal consistency

The reliability or internal consistency of the Full Scale IQ scores can be calculated through the Cronbach alpha. This coefficient assesses the singularity or internal consistency of a scale. The higher the Cronbach alpha, the more reliable the scale. Since the Full Scale IQ score is a linear combination of underlying subtest scores, no parameters are violated. This would not be the case if Cronbach alpha were applied to the subtest scores. The test manual presents internal consistencies for the construction samples of 0.96 to 0.98. These are remarkably high and undoubtedly the result of not only an inherently reliable test but also a very high number of subjects and a great deal of variance in the samples. The internal consistency for the Full Scale IQ score for the paper-and-pencil pilot sample is 0.70. It is 0.80 for the computer version. Both of theses are much lower than the statistics from the construction sample. A much lower number of subjects is one reason but, more importantly, an extremely truncated range of scores and variance of this sample. The standard deviations for the student pilots are one-half of the those in the construction sample and the variance, therefore, would be one-quarter. Truncated distributions not only suppress univariate and

Figure 1
MAB Version Distributions



multivariate correlations but also the Cronbach alpha which is based upon an average of all possible variable correlations. It is interesting, though, that with the same variance as the paper-and-pencil version that the computerized version would be so much higher.

Both versions have adequate reliability. The computerized version is significantly better.

Distributions

Figure 1 presents the distribution of Full Scale IQs for the two samples. A five-point rolling average was employed to allow for easier comparison. Both distributions are relatively normal. They appear to have similar variance. They appear not to be skewed. They appear to have similar kurtosis.

There are no obvious differences in the shapes of the two distributions.

DISCUSSION

The current work has found that there are few if any differences between student pilot candidates taking a paper-and-pencil version of the MAB versus the Armstrong Laboratory's computerized version. There are no mean differences for any of the summary IQ scores. There are no mean differences for any of the Verbal subtests and there are only minor and counterbalancing differences on two of the five Performance subtests. Differences on Performance subtests were more likely given the graphic nature of the stimuli.

One and two factor solutions for general, verbal, and performance intelligences indicated good concordance across both versions of the test as well as against the original construction sample data. Indeed, there was some evidence that the computerized version behaves better for pilots than the paper-and-pencil version.

Reliability analysis indicates that both versions are reliable. Interestingly, the computerized version is actually more reliable for pilots than the paper-and-pencil version. This is a quite unexpected result. While sample sizes differed, the smaller sample of 135 is still large enough to produce stable results.

Finally, the plots of the two distributions of Full Scale IQ scores shows no evidence of major differences between the two administration methods. While minor differences may appear, no significant range, variance, skew, or kurtosis issues are apparent. This is particularly true given the type of use to which these tests are put.

Recommendations

Strictly speaking, no further comparison studies are necessary. There is little need given the lack of differences found in the current study.

If resources are available, a within subject design would be of value. Here a large sample, perhaps 100, would be given both versions of the test. This would allow for the traditional alternate forms coefficient statistic. Correlations between tests should approach reliability and be in the .80 to .90 range. The difficulty of such a study is recruiting subjects to take two 1.5 hour IQ tests. A study of similar design but based upon the results of the current paper would be to simply have the subjects take both forms of the Performance subtests. These are the tasks which are most prone to decrement given computerization.

REFERENCES

Carretta, T.R. and Ree, M.J. (1993). Basic Attributes Test: Psychometric equating of a computer-based test. International Journal of Aviation Psychology, 3, 189-201.

Flynn, C.F., Sipes, W.E., Grosenbach, M.J., and Ellsworth, J. (1994) Top Performer Survey: Computerized psychological assessment in aircrew. Aviation, Space, and Environmental Medicine, 65, 39-44.

Jackson, D. (1984). Multidimensional Aptitude Battery Manual. Port Huron: Research Psychologist Press.

King R. and Flynn CF. (1995). Defining and measuring the "Right Stuff": Neuropsychiatrically Enhanced Flight Screening (N-EFS). Aviation, Space, and Environmental Medicine, 66, 951-956.

Retzlaff, P. and Gibertini, M. (1988). The objective psychological testing of Air Force officers in pilot training. Aviation, Space, and Environmental Medicine, 59, 661-663.

Retzlaff, P. and Gibertini, M. (1987). Air Force pilot personality: Hard data on "The Right Stuff". Multivariate Behavioral Research, 22, 383-399.

Appendix A

Correlation Matrix of Paper-and-Pencil Version

	VERBAL	PERF	FULL	INFO	COMP	ARITH	SIM
VERBAL	1.0000						
PERF	0.3430	1.0000					
FULL	0.8068	0.8293	1.0000				
INFO	0.7151	0.2135	0.5570	1.0000			
COMP	0.5134	0.1147	0.3825	0.3785	1.0000		
ARITH	0.5259	0.1751	0.4221	0.0763	0.1884	1.0000	
SIM	0.6309	0.2883	0.5565	0.4768	0.3041	0.1449	1.0000
VOCAB	0.6876	0.1432	0.4999	0.5492	0.3434	0.2561	0.3595
DIGSYM	0.1726	0.5017	0.4138	0.0002	-0.0543	0.2384	0.1639
PIXCOMP	0.1929	0.4853	0.4229	0.2535	0.1093	-0.0582	0.1798
SPAT	0.2558	0.7009	0.5924	0.1924	0.1003	0.2307	0.2444
PIXARR	0.1815	0.5136	0.4264	0.1310	0.2136	0.0700	0.1634
OBJASS	0.2625	0.5916	0.5243	0.2253	0.1316	0.0926	0.2197

	VOCAB	DIGSYM	PIXCOMP	SPAT	PIXARR	OBJASS
VOCAB	1.0000					
DIGSYM	0.0259	1.0000				
PIXCOMP	0.0934	-0.0290	1.0000			
SPAT	0.0637	0.3791	0.3258	1.0000		
PIXARR	0.1081	0.1463	0.0880	0.2781	1.0000	
OBJASS	0.1468	0.1544	0.2667	0.4179	0.1934	1.0000

Note: Verbal, Performance, and Full Scale IQ's correlations are based on scaled scores, while raw scores are used for the subtest correlations. The Jackson intercorrelation matrix is available in the manual.

Appendix B

Correlation matrix of computer version

	VERBAL	PERF	FULL	INFO	COMP	ARITH	SIM
VERBAL	1.0000						
PERF	0.4101	1.0000					
FULL	0.8145	0.8611	1.0000				
INFO	0.7895	0.3585	0.6665	1.0000			
COMP	0.5866	0.3698	0.5624	0.4091	1.0000		
ARITH	0.6146	0.2626	0.5074	0.3039	0.2265	1.0000	
SIM	0.7255	0.3359	0.6183	0.4859	0.4366	0.3027	1.0000
VOCAB	0.7559	0.3581	0.6510	0.5966	0.4143	0.2424	0.5426
DIGSYM	0.2936	0.5412	0.4978	0.1528	0.1444	0.2970	0.1864
PIXCOMP	0.3647	0.6922	0.6416	0.3569	0.3340	0.0994	0.3113
SPAT	0.2749	0.6875	0.5858	0.2322	0.2300	0.1901	0.1977
PIXARR	0.3268	0.6697	0.6024	0.2829	0.2363	0.2484	0.2109
OBJASS	0.3293	0.7136	0.6347	0.2216	0.2247	0.2354	0.3214

	VOCAB	DIGSYM	PIXCOMP	SPAT	PIXARR	OBJASS
VOCAB	1.0000					
DIGSYM	0.1740	1.0000				
PIXCOMP	0.3969	0.1579	1.0000			
SPAT	0.1940	0.2632	0.3834	1.0000		
PIXARR	0.1841	0.3581	0.3526	0.2895	1.0000	
OBJASS	0.2112	0.3071	0.4115	0.4498	0.4052	1.0000